

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

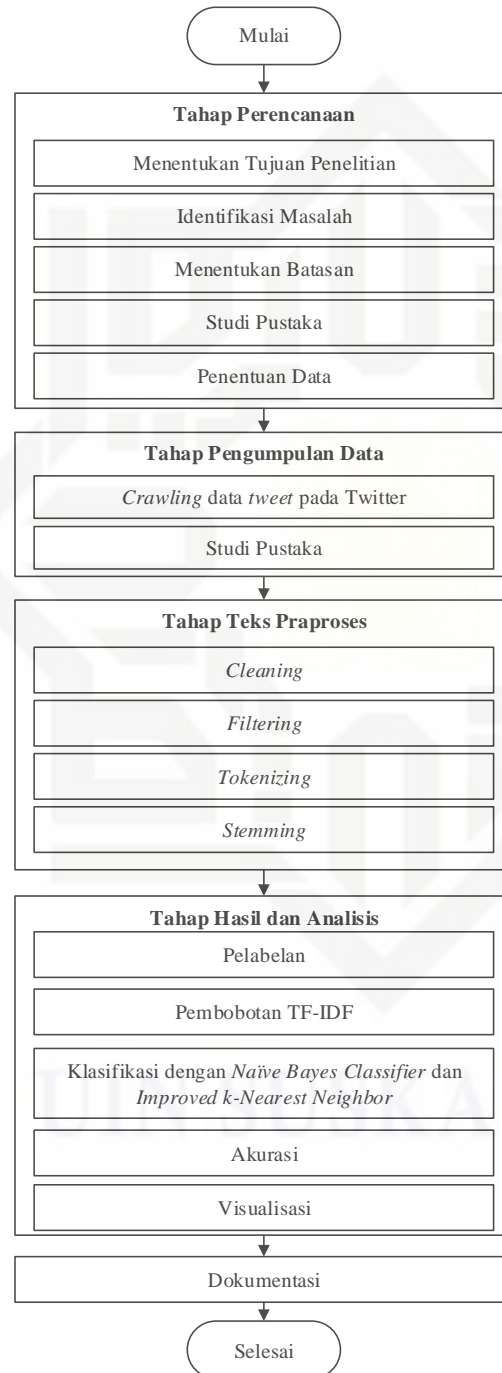
- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

BAB III

METODOLOGI PENELITIAN

Adapun langkah-langkah atau alur yang ditempuh dalam Tugas Akhir ini dapat dilihat pada Gambar 3.1.



Gambar 3.1 Metodologi Penelitian

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

3.1 Tahapan Perencanaan

Tahap perencanaan adalah tahapan yang harus direncanakan saat akan melakukan penelitian, data yang direncanakan adalah:

1. Penentuan Tujuan

Penentuan tujuan berfungsi untuk memperjelas kerangka tentang apa saja yang menjadi sasaran dari penelitian ini. Tujuan dari penelitian ini adalah melakukan klasifikasi terhadap opini masyarakat pada data yang didapat pada Twitter, Membandingkan Algoritma *Naïve Bayes Classifier* dan *Improved k-Nearest Neighbor* serta menghitung akurasi dari setiap algoritma.

2. Penentuan Masalah

Mengamati dan menemukan permasalahan yang terjadi pada Jasa Transportasi *online* dari opini masyarakat di Twitter.

3. Menentukan Batasan Masalah

Penentuan batasan masalah, agar penelitian tidak melebar terlalu luas dan tetap berfokus pada objek penelitian

4. Studi Pustaka

Bertujuan untuk mengetahui teori-teori apa yang akan digunakan untuk menyelesaikan permasalahan yang akan diteliti, serta mendapatkan dasar-dasar referensi yang kuat bagi peneliti untuk menganalisa opini masyarakat terhadap jasa transportasi *online* pada twitter menggunakan algoritma *Naïve Bayes Classifier* dan *Improved k-Nearest Neighbor*.

5. Penentuan data

Dataset yang digunakan dalam penelitian ini adalah data yang didapat pada setiap *tweet* dari *crawling* twitter. Adapun *tweet* yang digunakan adalah setiap *tweet* yang mengarah pada transportasi *online* yaitu @gojekindonesia, @UberIDN dan @GrabID.

Instrumen penelitian yang digunakan untuk melakukan eksperimen ini sebagai berikut.

- a. Sistem operasi yang digunakan adalah Windows 10 Enterprise 64-bit.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

- b. Python 2.7.
- c. Microsoft Excel 2013

3.2 Tahap Pengumpulan Data

Adapun proses yang terdapat pada tahap pengumpulan data adalah sebagai berikut:

1. Pengumpulan data twitter (*Crawling*)

Dalam pengumpulan data, peneliti mengambil data twitter yaitu data *tweet* yang mengarah pada akun jasa transportasi *online* yaitu @gojekindonesia, @UberIDN dan @GrabID.

2. Studi Pustaka

Studi literatur yaitu dasar teori yang digunakan dalam penelitian untuk menyelesaikan permasalahan dan merupakan referensi yang kuat dalam melakukan analisa.

3.3 Tahap Text Preprocessing

Adapun yang dilakukan pada tahap *Text Preprocessing* adalah sebagai berikut:

1. *Cleaning*

Proses *cleaning* yang dilakukan yaitu merubah teks menjadi huruf kecil (*case folding*), menghapus karakter selain huruf, menghapus *username* pengguna atau *mentions* (@), menghapus *hashtag* (#), menghapus URL atau *link* dari setiap *tweet*, serta menghapus *tweet* yang sama atau *retweets*.

2. *Filtering*

Pada tahapan *filtering*, yaitu pembuangan kata-kata tidak penting dari hasil token. Selain juga dilakukan penghapusan tanda baca dan *stopword*. *Stopwords* yang digunakan yaitu *stopword* Tala (Tala, 2003). *Stopword* didefinisikan sebagai sejumlah kata yang tidak berhubungan (*irrelevant*) dengan subyek utama yang dimaksud, meskipun kata tersebut memiliki frekuensi muncul yang tinggi dalam kata (Setiawan, *et al.*, 2013).

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

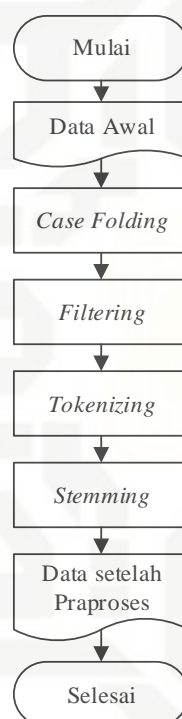
3. *Stemming*

Pada tahapan *stemming*, yaitu pengubahan kata berimbuhan menjadi kata dasar. *Stemming* yang digunakan yaitu algoritma Nazief dan Adriani.

4. *Tokenizing*

Pada tahapan *tokenizing*, setiap kata akan dipisahkan berdasarkan spasi yang ditemukan.

Alur pada pada proses *text preprocessing* pada dokumen atau *tweet* dapat dilihat pada Gambar 3.2.



Gambar 3.2 Alur *Text Preprocessing*

3.4 Tahap Hasil dan Analisis

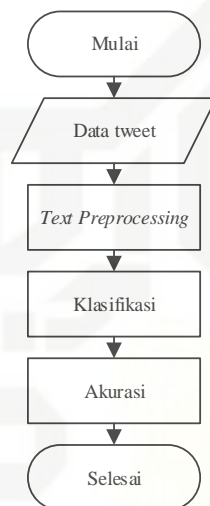
Adapun tahapan *text mining* dapat dilihat pada Gambar 3.3. Sedangkan, langkah-langkah yang dilakukan pada tahap hasil dan analisis adalah sebagai berikut:

1. Pembobotan

Tahap pembobotan adalah tahap pemberian bobot pada setiap kata dengan menggunakan Menghitung *Term Frequency* dan *Inverse Document Frequency*.

2. Analisis Klasifikasi

Setelah tahap sebelumnya selesai, dilanjutkan dengan melakukan klasifikasi dengan menggunakan algoritma *Naïve Bayes Classifier* dan *Improved k-Nearest Neighbor*. Setelah dilakukan perhitungan dengan menggunakan kedua algoritma, selanjutnya dilakukan perhitungan akurasi dari kedua algoritma. Data yang telah diklasifikasi dilakukan analisa pada setiap algoritma.



Gambar 3.3 Alur *Text Mining*

3.4.1 Contoh Perhitungan *Naïve Bayes Classifier* (NBC)

Tahapan awal pada proses NBC yaitu dengan menghitung probabilitas masing-masing kelas terhadap keseluruhan data latih. Probabilitas disimbolkan sebagai p.

$$p(Kelas_i) = \frac{Jumlah\ Kelas_i}{Jumlah\ data\ latih} \dots\dots\dots (3.1)$$

Berikut adalah hasil hitungan pada setiap kelas, terlihat pada Tabel 3.1.

Tabel 3.1 Probabilitas Kelas pada NBC

Kelas	Probabilitas Kelas
Netral	$p(Netral) = \frac{4}{15} = 0.267$
Negatif	$p(Negatif) = \frac{8}{15} = 0.533$
Positif	$p(Negatif) = \frac{3}{15} = 0.200$

Setelah menentukan probabilitas kelas, langkah selanjutnya dengan menentukan nilai n , nc , p dan m untuk setiap term yang terdapat pada data uji. Dari nilai-nilai yang telah diperoleh dan dengan menggunakan persamaan 2.5 diperoleh perhitungan sebagai berikut:

$$P(\text{Netral}|\text{stasiun}) = \frac{0 + 10.0,267}{30 + 10} = 0.067$$

$$P(\text{Netral}|\text{ui}) = \frac{0 + 10.0,267}{30 + 10} = 0.067$$

$$P(\text{Positif}|\text{rugi}) = \frac{0 + 10.0,200}{21 + 10} = 0.065$$

Berikut adalah nilai-nilai pada pada kelas Netral yang dapat dilihat pada Tabel 3.2.

Tabel 3.2 Nilai Kelas Netral

Netral	N	nc	p	m	hasil
stasiun	30	0	0.267	10	0.067
ui	30	0	0.267	10	0.067
titik	30	0	0.267	10	0.067
jemput	30	0	0.267	10	0.067
tolong	30	0	0.267	10	0.067
gojek	30	3	0.267	10	0.142
komunikasi	30	0	0.267	10	0.067
kelola	30	0	0.267	10	0.067
kmi	30	0	0.267	10	0.067
rugi	30	0	0.267	10	0.067

Untuk nilai pada kelas negatif dapat dilihat pada Tabel 3.3.

Tabel 3.3 Nilai Kelas Negatif

term	n	nc	p	m	hasil
stasiun	92	0	0.533	10	0.052
ui	92	0	0.533	10	0.052
titik	92	0	0.533	10	0.052
jemput	92	0	0.533	10	0.052
tolong	92	0	0.533	10	0.052
gojek	92	3	0.533	10	0.082
komunikasi	92	0	0.533	10	0.052
kelola	92	0	0.533	10	0.052

Tabel 3.3 Nilai Kelas Negatif (Lanjutan)

term	n	nc	p	m	hasil
kmi	92	0	0.533	10	0.052
rugi	92	0	0.533	10	0.052

Sedangkan, pada kelas positif dapat dilihat pada Tabel 3.4.

Tabel 3.4 Nilai Kelas Positif

Positif	n	nc	p	m	hasil
stasiun	21	0	0.200	10	0.065
ui	21	0	0.200	10	0.065
titik	21	0	0.200	10	0.065
jemput	21	0	0.200	10	0.065
tolong	21	0	0.200	10	0.065
gojek	21	3	0.200	10	0.161
komunikasi	21	0	0.200	10	0.065
kelola	21	0	0.200	10	0.065
kmi	21	0	0.200	10	0.065
rugi	21	0	0.200	10	0.065

Dengan menggunakan persamaan 2.6, yaitu mencari nilai maksimal dari hasil perkalian nilai probabilitas dan nilai P untuk setiap kelas, diperoleh hasil sebagai berikut:

$$V(\text{Netral}) = 0.267 \times 0.067 \times 0.067 \times \dots \times 0.067 = 9.84E - 13$$

$$V(\text{Negatif}) = 0.533 \times 0.052 \times 0.052 \times \dots \times 0.052 = 1.27E - 13$$

$$V(\text{Positif}) = 0.200 \times 0.065 \times 0.065 \times \dots \times 0.065 = 6.24E - 13$$

$$V_{nb} = \text{argmax} (v(\text{Netral})|v(\text{Negatif})|v(\text{Positif}))$$

$$V_{nb} = \text{argmax} (9.84E-13|1.27E-13|6.24E-13)$$

$$V_{nb} = 9.84E-13$$

Nilai maksimal yang diperoleh yaitu 9.84E-13, nilai v untuk kelas Netral. Jadi, kesimpulan yang diperoleh adalah data uji D16 termasuk kelas Netral.

3.4.2 Contoh Perhitungan *Improved k-Nearest Neighbor*

Tahapan awal yaitu menghitung bobot dari setiap dokumen. Berikut adalah *term frequency* pada dokumen dapat dilihat pada Tabel 3.5.

Tabel 3.5 *Term Frequency* (TF)

<i>Terms</i>	TF						
	D1	D2	D3	D4	D5	...	D16
	netral	netral	netral	netral	negatif	...	negatif
admin	1	0	0	0	0	...	0
ampe	0	0	0	0	0	...	0
aplikasi	0	0	0	0	0	...	0
...
yni	0	0	1	0	0	...	0

Selanjutnya, menghitung *Inverse Document Frequency* (IDF) didapat hasil seperti pada Tabel 3.6.

Tabel 3.6 *Inverse Document Frequency* (IDF) dan Wdt

<i>Terms</i>	df	idf=log (N/df)	Wdt = idf * tf					
			D1	D2	D3	D4	...	D16
			netral	netral	netral	netral	...	negatif
admin	1	1.204	1.204	0.000	0.000	0.000	...	0.000
ampe	1	1.204	0.000	0.000	0.000	0.000	...	0.000
aplikasi	1	1.204	0.000	0.000	0.000	0.000	...	0.000
...
yni	1	1.204	0.000	0.000	1.204	0.000	...	0.000

Setelah menghitung IDF pada dokumen, selanjutnya adalah menghitung panjang vektor dokumen. Didapat hasil seperti pada Tabel 3.7.

Tabel 3.7 Panjang Vektor

<i>Terms</i>	Panjang Vektor						
	D1	D2	D3	D4	D5	...	D16
	netral	netral	netral	netral	negatif	...	negatif
admin	1.450	0.000	0.000	0.000	0.000	...	0.000
ampe	0.000	0.000	0.000	0.000	0.000	...	0.000
aplikasi	0.000	0.000	0.000	0.000	0.000	...	0.000
...
yni	0.000	0.000	1.450	0.000	0.000	...	0.000
Total	5.981	3.791	8.397	21.477	18.214	...	13.578
Panjang Vektor	2.446	1.947	2.898	4.634	4.268	...	3.685

Setelah didapatkan panjang vektor, selanjutnya menghitung kedekatan antar data. Adapun hasilnya sebagai berikut:

$$D1(Netral) = \frac{0}{3.684 * 2.445} = 0$$

$$D2(Netral) = \frac{0}{3.684 * 2.445} = 0$$

.....

$$D15(Positif) = \frac{0}{3.684 * 1.028} = 0$$

Adapun hasil lengkap *cosine similarity* dapat dilihat pada Tabel 3.8.

Tabel 3.8 *Cosine Similarity*

No	Dokumen	Kelas	Kedekatan
1	D1	Netral	0.000
2	D2	Netral	0.000
3	D3	Netral	0.000
4	D4	Netral	0.000
5	D5	Negatif	0.000
6	D6	Negatif	0.000
7	D7	Negatif	0.000
8	D8	Negatif	0.000
9	D9	Negatif	0.087
10	D10	Negatif	0.034
11	D11	Negatif	0.000
12	D12	Negatif	0.000
13	D13	Positif	0.000
14	D14	Positif	0.000
15	D15	Positif	0.000

Lakukan pengurutan dari nilai yang tertinggi didapat hasil seperti pada Tabel 3.9.

Tabel 3.9 Kedekatan dengan Urutan dari Tertinggi

No	Dokumen	Kelas	kedekatan
1	D9	Negatif	0.087
2	D10	Negatif	0.034
3	D1	Netral	0.000
4	D2	Netral	0.000
5	D3	Netral	0.000
6	D4	Netral	0.000
7	D5	Negatif	0.000
8	D6	Negatif	0.000

Tabel 3.10 Kedekatan dengan Urutan dari Tertinggi (Lanjutan)

No	Dokumen	Kelas	kedekatan
9	D7	Negatif	0.000
10	D8	Negatif	0.000
11	D11	Negatif	0.000
12	D12	Negatif	0.000
13	D13	Positif	0.000
14	D14	Positif	0.000
15	D15	Positif	0.000

Dengan nilai $k = 3$. Didapatkan hasil yaitu data uji D16 mendapatkan kelas Negatif jika *k-Nearest Neighbor* cukup sampai tahap ini. Namun, dengan menggunakan *Improved k-Nearest Neighbor*, pertama dilakukan perhitungan jumlah data latih tiap kategori atau kelas seperti pada Tabel 3.10.

Tabel 3.11 Jumlah Data Latih

Data Latih			
Netral	Negatif	Positif	Jumlah
4	8	3	15

Kemudian dilakukan perhitungan nilai k baru pada masing-masing kelas. Perhitungan nilai n atau nilai k baru dihitung dengan menggunakan persamaan 2.7 yang dapat dilihat sebagai berikut:

$$n(\text{Netral}) = \frac{3 \times 4}{8} = \frac{12}{8} = 1.5 = 2$$

$$n(\text{Negatif}) = \frac{3 \times 8}{8} = \frac{12}{8} = 3$$

$$n(\text{Positif}) = \frac{3 \times 3}{8} = \frac{9}{8} = 1.12 = 1$$

Nilai dari k baru pada kelas netral, negatif dan positif dapat dilihat pada Tabel 3.11.

Tabel 3.12 Nilai n (k Baru)

Nilai k	n (k baru)		
	n Netral	n Negatif	n Positif
3	2	3	1

Tahapan selanjutnya yaitu dengan menghitung probabilitas pada masing-masing kelas dengan nilai k baru yang telah didapatkan dengan menggunakan persamaan 2.8. Perhitungannya probabilitas dengan nilai k baru pada tiap kelas dapat dilihat lebih jelas dibawah ini:

$$P(x, c_{netral}) = \frac{(0.087 \times 0) + (0.034 \times 0)}{(0.087 + 0.034)} = \frac{0}{0.121} = 0$$

$$P(x, c_{neg}) = \frac{(0.087 \times 1) + (0.034 \times 1) + (0.000 \times 0)}{0.087 + 0.034 + 0.000} = \frac{0.121}{0.121} = 1$$

$$P(x, c_{pos}) = \frac{0.087 \times 0}{0.087} = \frac{0}{0.087} = 0$$

Hasil probabilitas pada setiap kategori atau kelas juga dapat dilihat pada Tabel 3.12.

Berdasarkan hasil tersebut, probabilitas tertinggi didapat pada kelas Negatif. Sehingga dengan menggunakan *Improved k-Nearest Neighbor* data atau *tweet D16* diprediksi berada pada kelas Negatif.

Tabel 3.13 Probabilitas Kelas *Improved k-Nearest Neighbor*

k Awal	n (k Baru)		
	p Netral	p Negatif	p Positif
3	0	1	0

3. Akurasi

Pada tahapan ini dihitung akurasi pada kedua algoritma yaitu *Naïve Bayes Classifier* dan *Improved k-Nearest Neighbor* untuk melihat performa kedua algoritma.

4. Visualisasi

Tahapan ini yaitu akan membuat visualisasi kata dari dataset berdasarkan frekuensi kata yang muncul sesuai dengan kelas masing-masing yaitu netral, positif dan negatif

3.5 Tahap dokumentasi

Membuat dokumentasi hasil penelitian laporan tugas akhir Tahapan ini merupakan tahapan pengerjaan laporan. Hasil akhir pada tahapan ini berupa dokumentasi laporan tugas akhir.